

MESTRADO

**MÉTODOS QUANTITATIVOS PARA A DECISÃO
ECONÓMICA E EMPRESARIAL**

TRABALHO FINAL DE MESTRADO

PROJETO

**ANÁLISE DA EVOLUÇÃO DAS EMPRESAS POR
SEPARAÇÃO DE OBSERVADORES**

CRISTINA FILIPCIUC

NOVEMBRO - 2020

MESTRADO EM
MÉTODOS QUANTITATIVOS PARA A DECISÃO
ECONÓMICO E EMPRESARIAL

TRABALHO FINAL DE MESTRADO
PROJETO

ANÁLISE DA EVOLUÇÃO DAS EMPRESAS POR
SEPARAÇÃO DE OBSERVADORES

CRISTINA FILIPCIUC

ORIENTADORES:

PROF. CARLOS MANUEL JORGE DA COSTA

PROF. JOÃO CARLOS CAETANO DE FREITAS PIRES DA CRUZ

NOVEMBRO - 2020



LISBON
**SCHOOL OF
ECONOMICS &
MANAGEMENT**
UNIVERSIDADE DE LISBOA

GLOSSÁRIO

CIK – Central Index Key.

SEC – U.S. Securities and Exchange Commission.

SQL – Structured Query Language.

TLC – Teorema Limite Central.

WWW – World Wide Web.

RESUMO

O uso de dados de séries temporais na modelação de redes financeiras e económicas desafiam algumas suposições estatísticas tradicionais, como a aplicação do Teorema de Limite Central (TLC). No entanto, o recurso aos pressupostos provenientes da Física foi possível resolver algumas das limitações abordadas ao longo do documento. O trabalho desenvolvido baseia-se na aplicação de algoritmos de separação de observadores desenvolvidos pela Closer, cujo problema tem sido abordado desde há alguns anos, permitindo assim resolver os problemas associados à aplicação do TLC. Estes algoritmos baseiam-se na geometria diferencial e relatividade, que foram aplicados em séries de ações das empresas do mercado americano retiradas em escalas de tempo diversas, reportando no final os resultados obtidos em termos de transformação das distribuições vistas por cada um dos observadores.

Palavras Chave: Empresas americanas, Teorema do Limite Central, Teoria dos Grafos, Redes Complexas, Axiomas de Kolmogorov, Lei da Potência, Lévy e Física.

ABSTRACT

The handling of time-series data in modeling financial and economic networks challenges some traditional statistical assumptions, such as application of the Central Limit Theorem. However, using the assumptions of Physics it was possible to understand some of the limitations of the model. The work developed is set up on the application of observer separation algorithms developed by Closer Consulting, whose problem has addressed for some years, which solves the infinite variation limitation. These algorithms based on differential geometry and relativity, which was applied to series of shares of companies in the American market taken at different time scales, reporting, in the end, the results obtained in terms of transforming the distributions seen by each of the observers.

Keywords: American companies, Central limit theorem, Graph theory, Complex Networks, Kolmogorov axioms, Power Law, Lévy, and Physics.

ÍNDICE

Glossário	i
Resumo	ii
Abstract.....	iii
Índice	iv
Índice de Figuras	vi
Índice de Tabelas	vii
Agradecimentos	viii
1. Introdução	1
2. Revisão da Literatura.....	3
2.1 Definição de Economia	3
2.2 Trocas Económicas.....	3
2.3 Teoria dos grafos	4
2.4 Economia como Rede Complexa	6
2.5 Economia vista como Física	8
2.6 Modelo de agentes	11
3. O Modelo.....	15
3.1 Estudar o futuro	16
3.2 Definição do Problema	16
3.3 Agente de referência.....	19
3.4 Descrição do algoritmo.....	19
4. Tratamento Empírico.....	21
4.1 Origem da Base de Dados	21
4.2 Regras no tratamento dos dados	21
4.3 Dificuldades encontradas.....	22
4.4 Amostra final	23

5.	Resultados.....	24
6.	Conclusão	36
	Referências Bibliográficas.....	38
	ANEXOS	40

ÍNDICE DE FIGURAS

Figura 1 – Representação de possíveis exemplos de grafos.....	5
Figura 2 – (Sociedade - Rede Complexa de Relações, 2020)	6
Figura 3 – Distribuição Normal.....	7
Figura 4 – Exemplo de redes aleatórias e livre de escala.	7
Figura 5 – Representação expansão do espaço económico.	18
Figura 6 – Distribuição dos ativos para o último trimestre de 2019.....	24
Figura 7 – Distribuição dos passivos para o último trimestre de 2019.....	24
Figura 8 – Distribuição dos ativos ao longo do histórico.....	25
Figura 9 – Distribuição dos passivos ao longo do histórico.....	25
Figura 10 – Estatísticas obtidas para ativos (1) e passivos (2) dos dados originais.	28
Figura 11 – Estatísticas obtidas para ativos (1) e passivos (2) com dados relativizados.	31
Figura 12 – Estatísticas obtidas para ativos (1) e passivos (2) depois da aplicação do algoritmo.....	34
Figura 13 – Distribuição dos ativos após a aplicação do algoritmo.	35
Figura 14 – Distribuição dos passivos após a aplicação do algoritmo.	35

ÍNDICE DE TABELAS

Tabela I – Valores discriminados das estatísticas obtidas para os valores originais dos ativos.....	26
Tabela II - Valores discriminados das estatísticas obtidas para os valores originais dos passivos.....	27
Tabela III - Valores discriminados das estatísticas obtidas para os valores dos ativos relativizados.....	29
Tabela IV - Valores discriminados das estatísticas obtidas para os valores dos passivos relativizados.....	30
Tabela V - Valores discriminados das estatísticas obtidas para os valores dos ativos após a aplicação do algoritmo.....	32
Tabela VI - Valores discriminados das estatísticas obtidas para os valores dos passivos após a aplicação do algoritmo.	33
Tabela VII – Base de dados utilizada na modelação.	40

AGRADECIMENTOS

Em primeiro lugar, um especial agradecimento ao João Pires da Cruz e ao Professor Doutor Carlos Costa, pela paciência e por terem acreditado neste projeto.

Em segundo, agradecer ao ISEG, que sempre foi a minha segunda casa durante todo o meu percurso académico e por me ter cruzado com pessoas que levei comigo para a vida: Inês Couto, Dora Roxo, João Gil Delgado, Diogo Coelho, Carolina Marques, Beatriz Silva, Mafalda Fernandes, Marta Fernandes, Patrícia Ferreira e Catarina Lopes. Também deixo um agradecimento especial à Helena Faria e à Ana Rita Reis.

Por último, agradecer aos meus pais, sem os quais nada disto era possível e, ao Rafael Oliveira, pelo seu apoio incondicional e paciência infinita.

1. INTRODUÇÃO

O uso de séries temporais no âmbito da Economia e Finanças tem sido recorrente nas últimas décadas. Embora estas armazenem o traço quantitativo da atividade do ser humano durante décadas, desafiam algumas suposições estatísticas tradicionais, por exemplo, uma vez que as séries temporais observadas apresentam variâncias infinitas isso impossibilita a aplicação do TLC, bem como a suposição da correlação entre os agentes e económicos. Neste contexto, o trabalho de investigação aqui apresentado parte de uma questão relevante: como transformar distribuições com variância infinita numa variância finita?

Propõe-se como objetivo perceber em que contexto existem distribuições com variância infinita aquando o uso de séries temporais, assim como as possíveis soluções que permitam efetuar a transformação da variância infinita para variância finita. Nesse sentido, é utilizado um modelo desenvolvido pela Closer e aplicado a dados com ativos e passivos de empresas do mercado americano, de forma a obter uma distribuição normal, superando assim as limitações abordadas ao longo do documento.

Neste trabalho, assume-se que a Economia é constituída por agentes que efetuam trocas económicas, sendo importante compreender como é que essas trocas de bens serviços e que pressupostos são utilizados. Aqui, o sistema económico será modelado como uma rede complexa, baseando-se no uso do modelo de agentes.

A estrutura do documento está dividida em cinco partes:

1. **Revisão da literatura:** para além da definição de Economia, é discutida a visão da Economia pelos físicos e como podemos contornar as limitações do uso de dados de séries económicas e financeiras;
2. **Proposta do modelo:** onde é explicado o algoritmo de transformação fundamentado a partir de um modelo simples inspirado pela “expansão do universo económico” (ou aquilo que os economistas chamam de “inflação”);
3. **Tratamento Empírico:** breve descrição da base de dados utilizada e quais as principais dificuldades encontradas aquando o tratamento da amostra de ativos

e passivos de empresas americanas entre último trimestre de 2014 e o último trimestre de 2019;

4. **Resultados:** apresentação dos resultados em termos de transformação das distribuições vistas por cada um dos observadores;
5. **Conclusões:** apresentação dos principais desafios e preocupações, bem como uma breve discussão do potencial futuro deste tipo de modelações.

2. REVISÃO DA LITERATURA

2.1 Definição de Economia

No âmbito deste trabalho, e face ao detalhe exigido por um modelo de agentes, vamos usar uma definição no âmbito de a Economia é a ciência social que estuda como a sociedade opta por atribuir os seus escassos recursos, que têm utilizações alternativas, para fornecer bens e serviços para consumo presente e futuro. (Hoag & Hoag, 2006). Sendo esta constituída por vários agentes económicos, torna-se difícil modelar o comportamento dos agentes do sistema sem a concretizar, pelo que optamos por usar uma definição equivalente de “a economia é um sistema onde os seres humanos são impelidos a trocar formas de trabalho realizadas em diferentes momentos no tempo para satisfazer uma quantidade infinita de necessidades” (da Cruz, 2018). Esta definição leva a assumir uma relação entre os recursos – que se referem ao passado e, logo, necessariamente finitos – e a produção/consumo – que se referem ao futuro e, logo, necessariamente infinitos. Esta relação induz à conclusão de que o produto é uma função dos recursos e, assim, este ecossistema de agentes pode ser definido como um conjunto de objetos multiplicativos que estão correlacionados entre si, correlação essa habitualmente designada por *troca económica*. Por outras palavras, cada um dos agentes usa os seus recursos para adquirir/entregar produção gerada nesse instante. Uma vez correlacionados, é importante compreender que ações tornam possíveis as trocas de bens e serviços entre os agentes.

2.2 Trocas Económicas

Podemos ir mais longe e questionar o porquê da ciência económica. Sendo o Homem um animal racional e social, é natural vivermos em comunidade e existir a necessidade de colaboração. Apesar de não ser impossível, é difícil um agente conseguir satisfazer todas as suas necessidades individualmente. A necessidade de colaboração leva sempre a trocas e com um número suficiente de trocas, um sistema emerge, ao qual damos o nome de *mercado* (Hoag & Hoag, 2006).

No contexto económico, para cada bem ou serviço, tem de ser definido um valor monetário, tendo por base o custo de produção, que por sua vez pressupõe o custo da matéria-prima e a respetiva mão de obra, ou o valor que o mercado está disposto a pagar.

Contudo, nada é infinito. E é aqui que nasce o termo de escassez, que salienta o facto de que os recursos são finitos.

Pensemos então na utilidade ou raridade de um determinado recurso: se algo for muito útil ou indispensável ao ser humano, esse recurso será mais cobiçado. Caso isto fosse verdade, um copo de água seria mais valioso que um diamante. Então e quanto à raridade? Podemos possuir a garrafa de vinho mais rara do mundo, mas se o próprio vinho estiver estragado, a mesma perderá todo o valor associado.

E será que podemos definir o dinheiro como quantidade reservada? O dinheiro é omnipresente no sistema económico, porém só os Bancos Centrais é que o podem emitir. Neste caso, podemos afirmar que a economia de um determinado país é um sistema fechado em equilíbrio macroscópico. Contudo, esta idealização de equilíbrio (Robbins, 1932) é quase impossível de se obter num mundo real, uma vez que num sistema acaba quase sempre por entrar moeda externa, que por sua vez leva à respetiva criação de dinheiro, fazendo com que deixe de existir a necessidade de recorrer aos Bancos Centrais.

2.3 Teoria dos grafos

Diariamente, inúmeras trocas económicas são efetuadas, onde o dinheiro de cada agente pode ser trocado quer por produtos quer por serviços. No Universo Económico, cada agente pode ser representado por um nodo e a respetiva troca por um vértice, unindo esses dois nodos (Figura 1 (a)); ao fim do dia, teremos algo parecido à Figura 1 (b). No entanto, nenhum agente é o centro do universo.

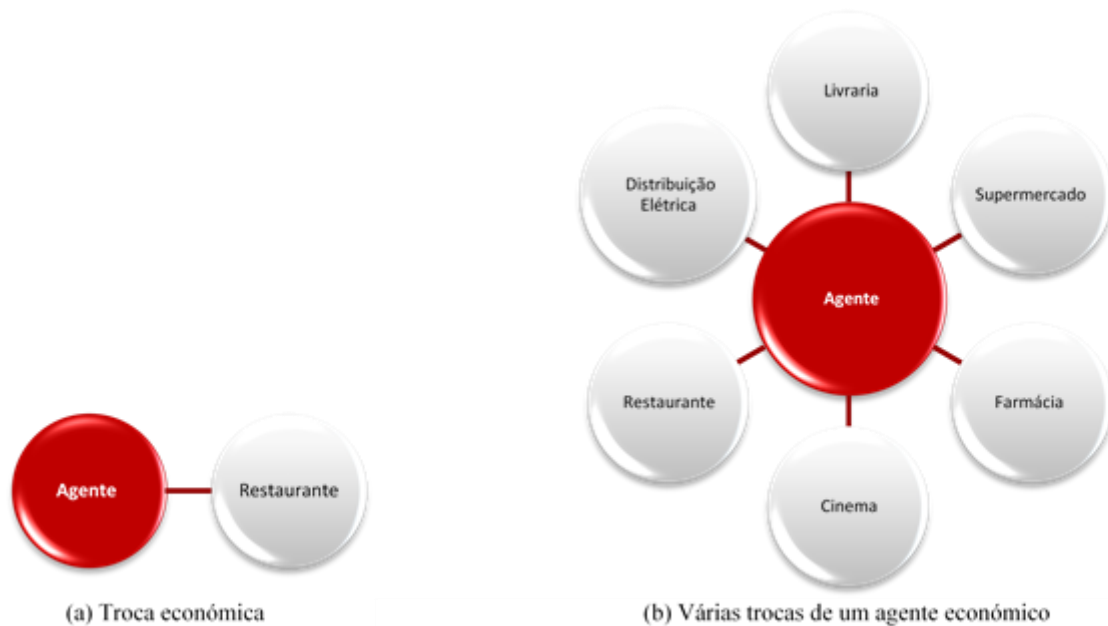


FIGURA 1 - Representação de possíveis exemplos de grafos.

Os exemplos observados anteriormente são denominados por *grafos*. A teoria dos grafos consiste num ramo da matemática dedicado ao estudo das relações entre elementos de um determinado conjunto (Biggs, Lloyd, & Wilson, 1986). É de esperar que quase todos os nodos façam trocas entre si, o que significa que esta rede de ligações económicas num determinado sistema será algo semelhante à Figura 2, que remete para uma possível ilustração de um grafo de diferentes trocas económicas de um sistema.

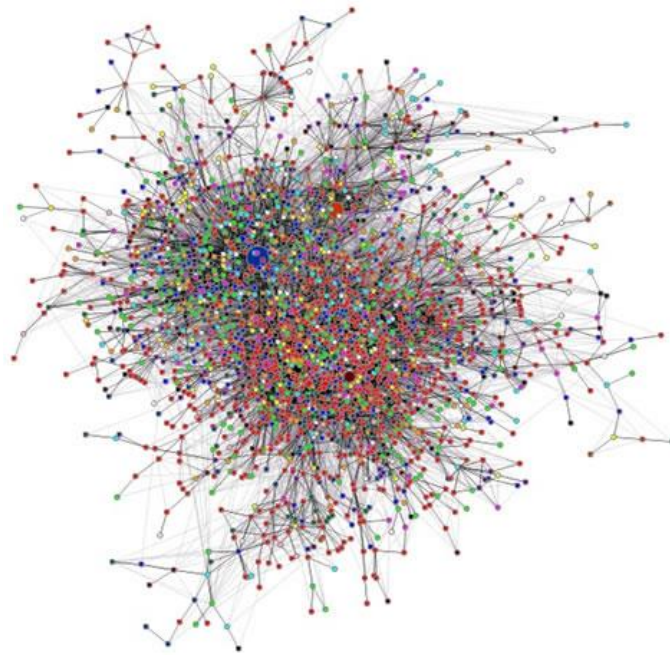


FIGURA 2 – (*Sociedade - Rede Complexa de Relações, 2020*)

Uma rede pode atingir rapidamente milhões ou biliões de nós e ligações, sendo possível caracterizá-la sem existir a necessidade da sua visualização.

2.4 Economia como Rede Complexa

Os grafos são formados por (e somente por) nós e ligações. Se for possível observar um comportamento emergente, isto é, um comportamento imprevisível, num grafo, denominamos o mesmo por *rede complexa* (Barabási & Réka, 2002). Em 1735, os grafos simples começaram a ser estudados por Euler, permitindo resolver o conhecido “Problema das Sete Pontes de Königsberg” (Euler, 1741).

As redes complexas começaram a ser cada vez mais admiradas, uma vez que se tornou perceptível a diversidade das suas aplicações, nomeadamente: WWW, redes sociais, redes neuronais, redes celulares e metabólicas, cadeias alimentares, cadeias de distribuição, entre outras (Barabási & Réka, 2002). Porém, a sua aplicabilidade pode ser alterada de acordo com a funcionalidade da rede, isto é, podemos recorrer a grafos orientados (separando as direções de cada ligação) ou a grafos bipartidos (uso de diferentes tipos de nodos, sendo que um nó só se pode ligar a nós do outro tipo) (Biggs,

Lloyd, & Wilson, 1986). É de esperar que as ligações que se formam em torno de nós tomem comportamentos “aleatórios” e sem padrão definido, uma vez que não podemos assumir a certeza dos acontecimentos. Por sua vez, estas implicam uma distribuição normal de graus, onde encontramos muitos nós perto da média e poucos nos extremos (Figura 3), ao qual designamos por redes aleatórias (*random networks*) (Figura 4 (a)).

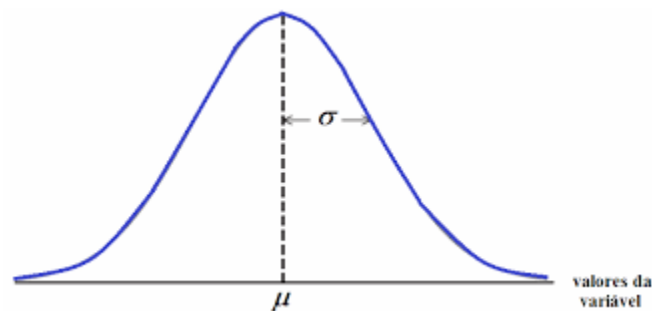
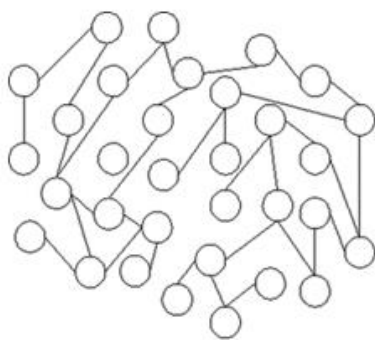
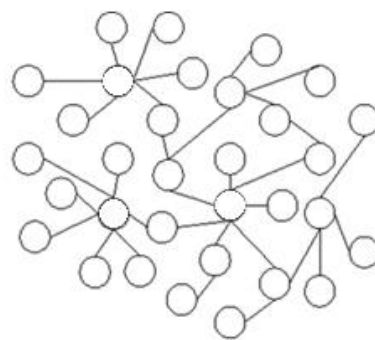


FIGURA 3 – Distribuição normal.

Porém, quando são efetuados estudos sobre as redes reais e é estudada a sua distribuição de graus, muitas vezes obtém-se uma lei de potência da forma $x^{-\alpha}$, onde $\alpha > 0$ e representa um expoente constante, isto é, ao modelar diferentes sistemas como uma rede complexa, obtemos para cada tipo de sistema um expoente característico (Pareto, 1964). Para este tipo de redes, dá-se o nome de livre de escala (*scale free*) (Figura 4 (b)).



(a) Rede aleatória.



(b) Rede *scale free*.

FIGURA 4 – Exemplo de redes aleatórias e livre de escala.

À semelhança das outras ciências sociais, a maioria das investigações realizadas nas áreas de finanças e economia apresentam uma abordagem na qual o comportamento de um sistema económico é explicado por algumas características chave, seja pela amplitude das flutuações de preços ou pela distribuição normal (Borland, Bouchaud, Muzy, & Zumbach, 2005). Estas características chave motivaram muitos investigadores a assumir distribuições de Lévy (Mantegna & Stanley, 1994). Esta suposição deriva da versão mais generalizada do TLC, em que é afirmado que a agregação de um número crescente de variáveis aleatórias converge para uma distribuição Lévy (Gnedenko & Kolmogorov, 1955). Caso estas variáveis aleatórias apresentem variâncias finitas, estamos perante uma distribuição normal, caso contrário, isto é, sendo a variância infinita, estaremos na presença de uma distribuição normal.

Numa abordagem epistemológica, o estudo destas distribuições, ignorando os argumentos mencionados anteriormente, torna-se possível ao criarmos uma função que se adapte a qualquer conjunto de dados empíricos com determinados parâmetros de ajuste, de modo que a função se adapte aos respetivos dados. Contudo, o uso de dados de séries temporais desafia algumas suposições estatísticas tradicionais. Embora estas armazenem o traço quantitativo da atividade do ser humano durante décadas, são bastante extensas, não seguem uma distribuição normal e não são estacionárias, o que significa que as suas propriedades estatísticas como a média e variância se alteram ao longo do tempo (Borland, Bouchaud, Muzy, & Zumbach, 2005). Neste caso, só poderíamos recorrer a esta abordagem caso os processos económicos fossem estacionários (Bouchaud & Potters, 2003, Sornette, 2003 e da Cruz & Lind, 2012).

2.5 Economia vista como Física

Alguns métodos derivados da Física e Estatística permitem explicar alguns dos fenómenos económicos, como a análise e previsão do mercado financeiro de ações e seus derivados (Mantegna & Stanley, 1999). O aumento do fluxo de informação disponibilizado a partir dos anos 80 motivou o estudo destes métodos, uma vez que começaram a ser adotados sistemas informáticos e com o aparecimento da Internet tornou-se possível a manutenção das bases de dados das empresas e respetivos governos.

Por norma, as técnicas utilizadas para este propósito envolvem processos estocásticos e incertezas.

Os conceitos como a mecânica newtoniana e a termodinâmica clássica (Mirowski, 1989) foram os primeiros a serem transferidos, originando, por sua vez, o conceito de equilíbrio das *forças* de procura e oferta, bem como foi feita uma tentativa de paralelismo com a energia livre (Smith & Foley, 2008). Outros modelos de equilíbrio também foram abordados, como o modelo Black-Scholes (Black & Scholes, 1973) para preços de derivados, que por sua vez foi muito explorado desde o início da década de 70, bem como a equação Fokker-Planck que acabou por ser utilizada na modelação dos rendimentos (Silva & Yakovenko, 2005 e Jagielski & Ryszard, 2013).

Todas estas técnicas assumem a premissa da conservação, uma vez que recorrem ao uso de *master equations*. Um sistema de *master equations* é um sistema de equações diferenciais que pressupõe uma modelação pela combinação probabilística de um número fixo de estados, onde a probabilidade de transição entre os estados é dada por uma matriz (Honerkamp, 2012). Num processo estocástico, as condições de equilíbrio presumem uma variável que seja preservada no tempo e que apresente uma propriedade da não variância no tempo (Kampen, 2006). Já em Física Estatística, quando um *conjunto microcanónico* (um conjunto de estados possíveis de um sistema fechado e isolado) assume a conservação da energia total, isso implica que o número total de microestados (cada um dos estados possíveis) é constante ao longo do tempo (Pires, Cruz, Rajaratnam, Beling, & Overstreet Jr., 2016).

Uma vez que as distribuições *normal* são observadas nas variáveis económicas, seria de esperar que os investigadores usassem as distribuições Lévy. O caso da distribuição normal, que representa um caso particular de uma distribuição de Lévy, foi o primeiro a ser considerado para a modelação de preços de ações Europeias pelo modelo Black-Scholes. No entanto, este modelo torna-se inconsistente uma vez que o comportamento dos preços das ações não segue uma distribuição normal. Para ultrapassar as imperfeições deste modelo, outros modelos mais sofisticados foram propostos desde as décadas de 1980 e 1990, que assumiam processos mais generalizados do que os processos Brownianos (Shoutens, 2003). Esses processos alternativos denominam-se por

processos de Lévy e as distribuições das probabilidades dos seus incrementos são infinitamente divisíveis, ou seja, uma variável aleatória seguindo essa distribuição pode ser decomposta numa soma de um número inteiro arbitrário de variáveis aleatórias independentemente distribuídas de forma idêntica (Shoutens, 2003).

Ainda assim, apesar dos avanços consideráveis na modelação de dados financeiros com processos de Lévy, os investigadores continuam a mostrar uma forte preferência pela classe particular de distribuições normais e, na verdade, existem boas razões para isso. Assumindo que as distribuições Lévy são boas representações das flutuações das variáveis económicas, um modelo baseado nelas é fechado quando se ajustam as distribuições aos dados empíricos escolhendo adequadamente os valores dos parâmetros (da Cruz & Lind, 2013) e é muito mais simples ajustar dois parâmetros (Normal) que quatro parâmetros (Lévy genérica) estando igualmente a laborar numa aproximação pouco rigorosa, uma vez que parte dos estudos estatísticos são desenvolvidos sobre a hipótese da normalidade.

Numa abordagem mais ontológica, ao modelar redes financeiras e económicas, as variáveis aleatórias são traduzidas em agentes. Tal como foi discutido anteriormente, o sistema económico pode ser encarado como um sistema de agentes que trocam de produto, mas os físicos vêem isso como uma quantidade fundamental que é comum ao trabalho, dinheiro ou salário. Tem de existir algo comum a todas as quantidades trocadas, caso contrário, não seriam trocas. Os físicos consideram essa quantidade fundamental como *energia económica*.

O Homem consome e produz quantidades finitas, sendo que esta suposição justifica o aparecimento para cada agente de um máximo de uma produção e o mínimo de consumo. Se um agente ultrapassa esses limites, ele deixa de poder consumir. Uma vez que o número de trocas é regulado apenas com os princípios de oferta e procura, o sistema pode levar um agente a exceder a sua fração máxima da quantidade total de trocas. Isto também é um fenómeno recorrente nos sistemas económicos, denominando-se tipicamente por casos de insolvência. O agente ao perder os elos de consumo, entra em colapso e leva à quebra de alguns elos de produção dos seus vizinhos, alterando também a fração máxima, levando também ao seu colapso. Consequentemente, cada colapso é

capaz de desencadear uma reação em cadeia originando um processo de ramificação. No contexto económico, uma reação em cadeia é denominada de “crise”, enquanto no contexto físico é normalmente chamada de “avalanche”. Ainda assim, é difícil contar o número de agentes em colapso numa rede real. O que é medido quando ocorre uma avalanche numa rede real é o número de ligações responsável por uma propriedade macroscópica do sistema (da Cruz & Lind, 2013).

Contudo, em Economia não se segue esse tipo de descrição do sistema (Robbins, 1932), uma vez que apresenta crescimento e é um sistema inflacionário onde a noção de equilíbrio não se relaciona com a preservação. Na verdade, em Economia, o equilíbrio surge como pressuposto e não como algo que precisa de ser verificado.

2.6 Modelo de agentes

Em vez de seguir uma modelação macroscópica, podemos recorrer a uma análise microscópica, isto é, dando um conjunto de regras a cada elemento de um conjunto de agentes (*agent-based modeling*) (Turrell, 2016), podemos simular um sistema de agentes e observar o comportamento emergente respetivo.

A macroeconomia é caracterizada pela existência de ciclos económicos. Na modelização macroeconómica clássica, as flutuações do crescimento do PIB resultam de choques exógenos (Gualdi, Tarzia, Zamponi, & Bouchaud, 2015). No mundo atual, as flutuações do PIB são provavelmente endógenas, ou seja, geradas pela própria economia. No entanto, estes modelos permitem tornar essas flutuações endógenas e, portanto, também permitem compreender as respetivas causas e que políticas podem afeta-las.

O modelo de agentes é adequado para modelar sistemas complexos, como a economia, nomeadamente sistemas económicos onde existe uma combinação de interações de diferentes agentes produzindo, por isso, resultados inesperados. Na economia, este tipo de modelo permite explicar como é que ocorrem os ciclos económicos e o surgimento das estatísticas observadas nos mercados financeiros (como é o caso das *fat tails*, cuja distribuição de probabilidade exige grande assimetria ou curtose quando comparadas com as distribuições normais ou exponenciais). Ainda assim, são um forte

complemento na modelação económica tradicional, que foi muito criticada após a Grande Recessão (Turrell, 2016). A Grande Recessão desafiou os processos de modelação económica, uma vez que mostrou que a economia é, de facto, muito complexa e que nem sempre assenta num equilíbrio estável.

De facto, os agentes representam consumidores numa economia, onde os comportamentos ou regras que esses agentes seguem dependem sobretudo das suas preferências. Os modelos podem ter diferentes tipos de agentes, empresas, trabalhadores e governos. Enquanto os trabalhadores podem apresentar diferentes níveis de produtividade, as empresas podem variar na respetiva dimensão.

Enquanto nas ciências naturais, os comportamentos dos agentes são muito mais restritos, na economia não é isso que se verifica. Nesse sentido, os modelos económicos baseados em agentes produzem *insights* em vez de previsões quantitativas (Turrell, 2016). Na verdade, são mais qualitativos do que quantitativos e, por isso, são bons para determinar possíveis cenários, em vez de afirmar aquilo que poderá realmente acontecer.

Uma importante aplicação destes modelos está relacionada com temas de saúde e epidemiológicos, isto é, no estudo da propagação de doenças. Ao compreender a dinâmica complexa das epidemias poderá, de facto, salvar milhões de vidas, uma vez que permitem recolher informações específicas de cada país, por exemplo: os dados geográficos, padrões de deslocamento diário, distribuição de idades e outras informações (Degli Atti, et al., 2008). De certa forma, é aquilo que se tem vindo a aplicar no nosso atual contexto pandémico.

No entanto, a aplicação destes modelos em contextos económicos e ciências sociais acarreta sempre vantagens e desvantagens da sua possível aplicação. Seguem-se abaixo as principais vantagens da utilização destes modelos (Turrell, 2016):

1. É possível observar um comportamento “emergente”, o que significa que o comportamento individual de cada um dos agentes é combinado para produzir comportamentos macroscópicos não observáveis microscopicamente;

2. Heterogeneidade. A modelação individual de cada agente permite explorar as consequências dos agentes serem heterogéneos, isto é, de certa forma nenhum agente é igual ao outro, sejam essas diferenças observadas ao nível do rendimento, das preferências, na educação ou na produtividade. A inclusão da heterogeneidade permite, nesse sentido, tornar a modelação muito mais rica em comportamento;
3. A formação de comportamentos realistas com base no comportamento observado pode ser um ponto forte neste tipo de modelos. Uma vez que as pessoas costumam usar heurísticas aquando a toma de decisões e que as mesmas nem sempre são totalmente racionais. Este tipo de comportamentos pode ser descrito coletivamente como “racionalidade limitada”;
4. Um dos maiores sucessos deste modelo na economia foi a explicação dos factos estilizados observados nos mercados de ativos. (Hong & Stein, 1999). Existe uma série de fenómenos observados empiricamente nos mercados de ativos que não são explicados pela teoria económica tradicional. As *fat tails*, que ocorrem em casos extremos, como grandes mudanças de preço, ocorrem com mais frequência do que seria esperado se uma distribuição normal fosse assumida;
5. As regras de probabilidade aplicadas a cada agente individualmente permitem explorar possíveis cenários, em vez de calcular como uma população se deve comportar em conjunto, como por exemplo nos estudos epidemiológicos;
6. Por último, estes modelos podem descrever sistemas complexos. Os sistemas complexos, tal como foi referido em 2., são caracterizados pela elevada interligação entre os agentes, tendo variáveis que podem mudar drasticamente e que poderão demonstrar auto-organização.

Do outro lado da moeda, podemos encontrar algumas desvantagens na aplicação destes modelos, uma vez que:

1. Uma das maiores vantagens destes modelos – a sua flexibilidade – também é apresentada como uma desvantagem. A escolha na construção destes modelos, em comparação com os modelos económicos tradicionais, implica que no passado existiram dificuldades na seleção dos componentes certos para o problema em estudo. Os resultados de uma determinada simulação podem variar drasticamente, sendo que deve existir um cuidado duplicado na escolha das premissas a usar;
2. Uma outra fraqueza assenta no facto deste tipo de modelos serem feitos sob medida, isto é, não são transversais na sua aplicabilidade.

3. O MODELO

A riqueza de um agente pressupõe a quantidade de recursos nele alocados (representada por x), que resulta da acumulação de trocas. Enquanto o produto é o aumento dos recursos alocados causado aquando trocas (representada por dx).

Reiterando uma vez mais a importância do produto, além da sua natureza específica, é importante estabelecer uma relação entre os recursos económicos a ele alocados e a sua contribuição no sistema. Aqui é observada uma diferença fundamental entre a Física e a Economia, onde a primeira apoia-se em leis de conservação e, a segunda, nos conceitos como riqueza e dívida, que aumentam proporcionalmente ao seu tamanho, contribuindo para o crescimento do sistema, não existindo conservação. Uma pessoa ao trabalhar, onde o seu esforço faz com que se produza a riqueza, fará com que exista uma expansão do próprio universo económico. Quando um agente económico produz (dx) em proporção ao que já possui (x), é possível assumir que:

$$dx = \beta x \Leftrightarrow \beta = \frac{dx}{x} \Rightarrow \beta = d \log x \quad (1)$$

Na equação apresentada acima (Equação 1), β representa a concretização matemática de um agente económico, isto é, representa a relação entre os recursos associados ao respetivo agente e aos que o próprio produz. Para além desta definição exata, é importante realçar que este representa um processo multiplicativo dx/x .

Aquando a interação ente os agentes, mesmo com ele próprio, provoca um crescimento de todos os intervenientes que nela participam. Este fenómeno é designado por *processo multiplicativo*, aos quais Piketty (Blanchett, Fournier, & Piketty, 2017) denomina por *choques multiplicativos aleatórios*, representa um mecanismo fundamental no aparecimento de distribuições que seguem uma lei de potência. Na verdade, todos os processos multiplicativos levam a uma distribuição com a forma de uma lei de potência (Sornette & Cont, 1997).

3.1 Estudar o futuro

Na tentativa de prever a evolução de um determinado sistema, surge o conceito de *probabilidade*. Em 1933, foi Andrey Kolmogorov que estabeleceu três axiomas que para além de unir as variadas interpretações de probabilidade, nomeadamente a clássica e a frequentista, estabelecem ainda o formalismo matemático de probabilidade. Estes axiomas são dados pelas seguintes fórmulas:

$$P(E) \in \mathbb{R}, P(E) \geq 0 \quad \forall E \in F \quad (2)$$

$$P(\Omega) = 1 \quad (3)$$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i) \quad (4)$$

Ora, para um acontecimento E temos um espaço de probabilidade definido por (Ω, Σ, P) onde Ω é o espaço de amostras, Σ o espaço de eventos e P a probabilidade de E .

A Economia sendo um sistema em constante expansão e dotado de correlação entre as partes, os recursos alocado a um agente acabam por gerar riqueza. Assim, ao normalizar uma variável económica para um determinado instante t_0 , em conjunto com a expansão da economia (quer pelo crescimento do negócio ou por inflação monetária) esta normalização deixa de ser válida para $t > t_0$. É possível concluir que a equação 4 deixa de ser verificada e, por isso, o conceito de probabilidade, qualquer que seja a sua definição ou caracterização, deixa de ser aplicável.

3.2 Definição do Problema

O facto de que uma variável económica, como a riqueza, o rendimento ou dívida, não apresentar uma distribuição normal, é muitas vezes ignorado nas análises económicas. As variáveis que apresentam variância infinita que consequentemente invalidam a aplicação do TLC. Também não são aprendíveis por processos mecânicos,

uma vez que o erro do processo de aprendizagem aumenta com o tamanho da amostra. Isto leva a concluir que os métodos de *Big Data* nunca serão robustos o suficiente para resolver este problema.

O trabalho aqui apresentado propõe uma métrica (uma medida de distância num espaço abstrato) que, sendo aplicada a dados com distribuição *fat tail*, terá de resultar numa distribuição quase normal onde, posteriormente, será medida a “Normalidade” dos respetivos dados usando a variância, curtose e simetria (permitindo passar de variância infinita para finita).

Sabendo que a distribuição de uma variável económica numa população tende a seguir uma lei de potência, pelo facto de esta apresentar uma variância infinita, qualquer extrapolação ou interpolação torna-se inútil, bem como as respetivas “previsões”. Seguindo uma distribuição *normal*, os valores extremos surgem com mais frequência do que seria expectável numa distribuição normal.

Resumindo:

1. É difícil modelar distribuições em lei de potência;
2. Um sistema económico pode ser modelado por uma rede complexa, contudo o conceito de rede complexa não é suficiente para um estudo rigoroso dos fenómenos presentes.

Uma rede complexa é constituída por nós e ligações, o que faz com que não exista “espaço circundante”, o que significa que toda a geometria dependesse do tempo, uma vez que o espaço económico é formado pelos agentes que dele fazem parte e, por consequência, evoluem em conjunto (Figura 5). No entanto, ao contrário de outros sistemas físicos, o espaço circundante (zona sombreada) não existe: o espaço económico é formado pelos agentes (nodos) e pelas respetivas ligações (vértices).

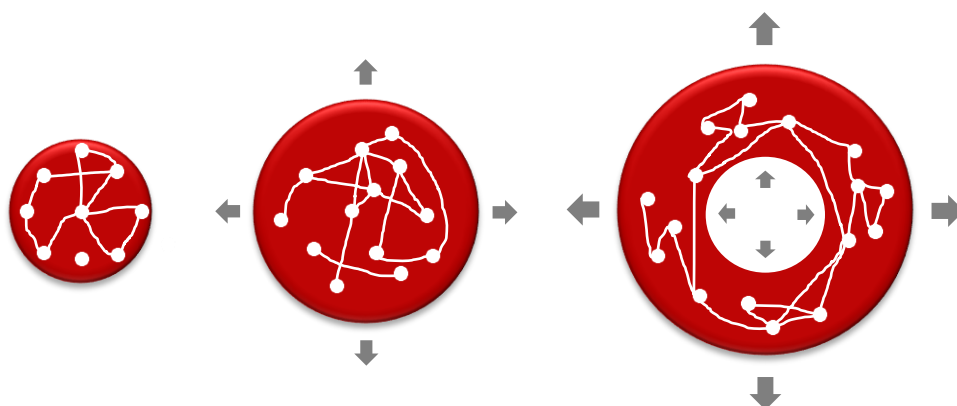


FIGURA 5 – Representação expansão do espaço económico.

No entanto, o espaço económico pode ser estudado através de algumas grandezas económicas, tais como os ativos e passivos. Ao longo do tempo e, graças às sucessivas interações entre os agentes, ocorre uma expansão quer dos próprios agentes quer do espaço onde estes se encontram. Contudo, os limites do espaço que, por sua vez, são definidos pelos *outliers*, refletem a realidade de algumas entidades no mundo financeiro: existem sempre entidades cujo valor da grandeza económica em estudo é mais alto e, acima desse valor, deixa de existir espaço económico. Ainda assim, existe uma zona vazia, que corresponde a um limite inferior abaixo do qual já não existem agentes e, nesse sentido, não pode existir espaço económico. A Figura 5 pode ser interpretada recorrendo ao conceito de zona de “economia ativa” que vai expandindo com a inflação e outros fatores.

Todas estas circunstâncias invocam os axiomas de Kolmogorov: quando, num determinado instante, ao estabelecer um espaço de probabilidade, no instante seguinte o universo aparenta ser diferente (uma vez que se alteram as riquezas, que podem aumentar ou diminuir ou, novos agentes são criados ou os antigos desaparecem), isto faz com que o axioma da unidade de medida deixe de ser respeitado. Em conjunto com os momentos infinitos das distribuições observadas num sistema económico, faz com que qualquer tipo de previsão deixe de ser válida.

3.3 *Agente de referência*

O modelo aqui proposto pressupõe a existência de um observador móvel para o qual o espaço económico é uniforme, isto é, este observador verá os restantes agentes uniformemente distribuídos em torno do próprio. Num determinado instante, o objetivo é procurar uma normalização que transforme a distribuição da variável económica numa distribuição aproximadamente normal. O problema é que quando medidos as variáveis económicas em unidades monetárias, estas representam um observador fixo. O valor numérico de uma nota de 10 euros será sempre 10 euros, embora o sistema económico, na sua expansão, se vá afastando desse referencial (“inflação”). O agente de referência é um agente, abstrato, que segue no tempo solidário com o sistema apesar da inflação, enquanto o observador fixo, o dinheiro, se vai afastando. Por outras palavras, enquanto o observador fixo vê um sistema em expansão e, logo, impossível de tratar estatisticamente, o observador de referência vê um sistema estacionário e, logo, passível de ser tratado dentro do domínio da estatística.

3.4 *Descrição do algoritmo*

Assim, estabeleceu-se um algoritmo de transformação que, aplicado a uma distribuição com variância infinita, permite transforma-la numa variância finita. Esta transformação é fundamentada a partir de um modelo simples inspirado pela “expansão do universo económico”.

Para poder normalizar uma variável económica, teremos de formar um novo entendimento de economia, onde existem apenas dois conceitos: agente e ligação. Nesse sentido, o nosso objetivo será encontrar agentes de referência que nos permitam transformar a respetiva população. Localmente (e relembrando a métrica proposta na equação 1), podemos escrever:

$$\beta = \frac{dx}{x} = d \log x = \log x - \log x_{ref} , \quad (5)$$

onde x_{ref} representa o nosso agente de referência: este pode ser qualquer agente existente ou a média da amostra ou qualquer outro valor dentro da mesma. O algoritmo a usar será composto pelos seguintes passos:

1. Calcular o logaritmo de cada agente:

Dada a importância do processo multiplicativo e a sua ligação com a distribuição em forma da lei de potência, será necessário olhar para a distribuição do logaritmo de cada agente, trazendo, assim, a vantagem de reduzir a variância da distribuição do mesmo.

2. Calcular a extensão da população:

Ao fazer diferença entre o agente máximo e mínimo da população, obtemos um fator de normalização a que designamos por fator c . Para calcular este fator, usamos a equação:

$$c = \frac{x_i(\text{ano actual}) - x_i(\text{ano anterior})}{x_i(\text{ano anterior})}, \quad (6)$$

onde x_i é o valor de ativos/passivos do agente mínimo escolhido.

3. Calcular a distância de cada agente a um agente de referência:

A escolha do agente em si não é relevante. Este passo permite uma abstração dos valores absolutos da população, permitindo assim olhar para a distância entre agentes. Note-se que este passo permite evitar o problema do que referência fixa que é a moeda cunhada.

4. Estudar a evolução das distâncias:

Este último passo pretende-se estudar a estatística em cada instante das distâncias ao observador de referência e o comportamento dos parâmetros estatísticos no tempo.

4. TRATAMENTO EMPÍRICO

4.1 Origem da Base de Dados

Foram analisados registos trimestrais de ativos e passivos de empresas americanas considerando o histórico desde 2º trimestre 2009 ao 1º trimestre de 2020. Toda a informação encontra-se disponível na SEC (SEC.GOV, s.d.), sendo uma base de dados pública e que sofre atualizações trimestrais.

4.2 Regras no tratamento dos dados

Para garantir o rigor e qualidade dos dados, foi preciso asseverar alguns pressupostos de modo a evitar duplicação dos valores de ativos e passivos para cada trimestre associado.

Primeiro, para tornar a leitura dos dados mais intuitiva, isto é, dado que a base de dados apresenta vários ficheiros em formato de texto (.txt) independentes para cada um dos trimestres, foi necessário transformar os respetivos ficheiros em tabelas e armazená-las numa única tabela.

Segundo, para importar e intercalar os respetivos dados, recorreu-se ao uso do Power Query (*add-in* do Excel). Feito o importe, utilizou-se a Power Pivot (outro *add-in* do Excel), permitindo assim armazenar todas as tabelas no modelo de dados do Excel, no qual posteriormente foram criadas relações entre as respetivas tabelas com base nos pressupostos do SQL.

Por fim, a tabela única deverá conter empresas que tenham registo de atividade em todos os trimestres do histórico analisado. Feito esse controlo, verificou-se que para as datas mais antigas, existiam poucas entidades registadas, isto é, existindo poucas empresas, a junção das tabelas com base nessas entidades teria apenas 49 registos para analisar. Nesse sentido, foram feitos vários testes de modo a conseguir obter pelo menos 2 000 agentes ao longo do período todo, sendo que tal resultado só foi obtido tomando como referência o último trimestre de 2014.

Em suma, a criação da tabela única teve em consideração os seguintes passos:

1. As entidades apresentadas no último trimestre de 2014 foram tidas como referência de modo a garantir que cada entidade esteja presente em todos os trimestres do período de análise;
2. Garantir que para cada par (empresa, trimestre) existam valores de ativos e passivos;
3. Separar a tabela inicial em duas: uma apenas para os registos de ativos e outra para passivos;
4. Para cada uma das tabelas, considerar o ADSH (EDGAR *accession number*) respetivo para cada registo e fazer o *merge* dos resultados numa outra tabela;
5. Com base dos resultados do *merge* obtido no passo anterior, associar os códigos de cada ADSH a um CIK respetivo. Obtém-se, assim, uma outra tabela com a lista dos CIK's para o histórico considerado;
6. Da tabela onde obtivemos as entidades que aparecem no período todo deste 2014T4, foi feito um *merge* com a tabela dos CIK's obtidos, de modo a obter a tabela final com os registos de ativos e passivos para cada entidade.

4.3 Dificuldades encontradas

Dada a especificidade da BD, foi tido em conta um cuidado adicional sempre que eram extraídos os valores dos passivos e ativos, uma vez que para um determinado trimestre, todas as entidades apresentam um registo histórico de ativos e passivos. Essa duplicação de valores pode refletir registos observados no trimestre que estamos a ter como referência e o trimestre anterior (ou até trimestres mais antigos). Esse registo acontece devido a sucessivas atualizações da própria BD que, por sua vez, permite visualizar a respetiva correção dos valores de um trimestre para o outro, sendo que apenas o último registo da matriz é considerado como o mais correto.

Para evitar esta duplicação, foi criada uma tabela mestre que permite visualizar os últimos registos para cada um dos trimestres onde, curiosamente, os valores contabilizados encontram-se sempre sobre a diagonal.

4.4 Amostra final

Inicialmente, a base de dados apresentava cerca de 612 000 registos indexados por número de identificação e respetivo trimestre, onde para cada registo estava associado um valor de ativo e passivo. Considerando as restrições mencionadas em 4.2, foram obtidos 2 065 agentes com registos desde o último trimestre de 2014, onde para cada par (empresa, trimestre) não existem registos duplicados. O ficheiro base pode ser consultado na Tabela VII apresentada nos Anexos.

5. RESULTADOS

A título de exemplo, nas Figura 6 e 7, temos as frequências das distribuições dos ativos e passivos em 2019 ainda sem a aplicação do algoritmo, ou seja, na presença de um observador fixo.

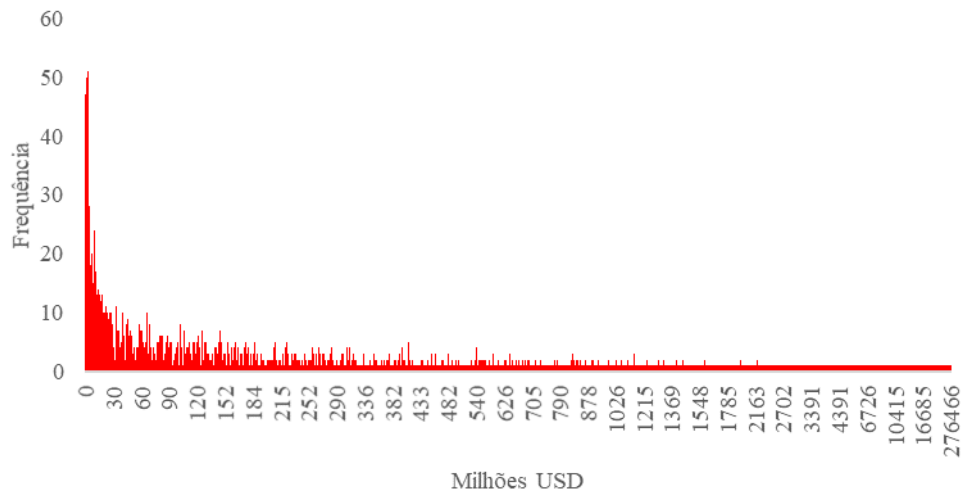


FIGURA 6 – Distribuição dos ativos para o último trimestre de 2019.

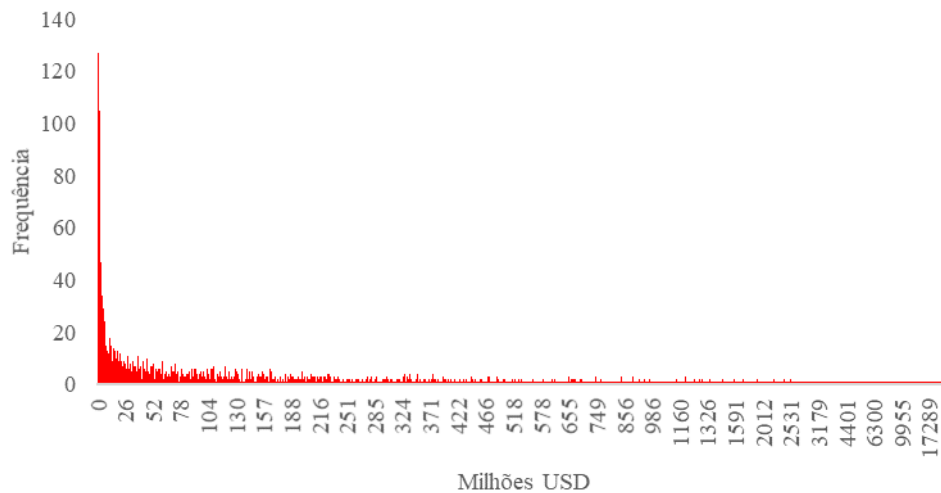


FIGURA 7 – Distribuição dos passivos para o último trimestre de 2019.

Quer para os ativos quer para os passivos observamos a esperada forma de uma lei de potência. Verificamos que existem uma concentração do número de entidades financeiras com os valores de passivos e ativos enviesados à direita.

Para o histórico todo os resultados obtidos foram semelhantes aos mencionados acima:

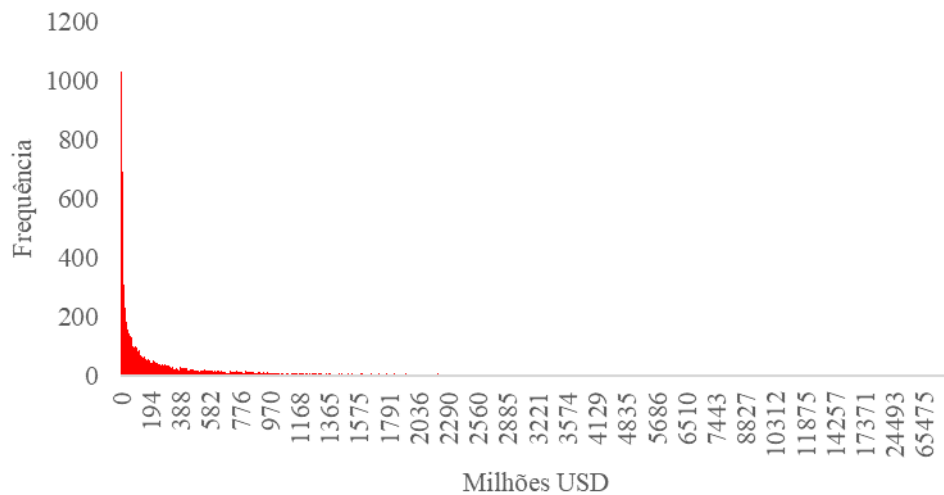


FIGURA 8 – Distribuição dos ativos ao longo do histórico.

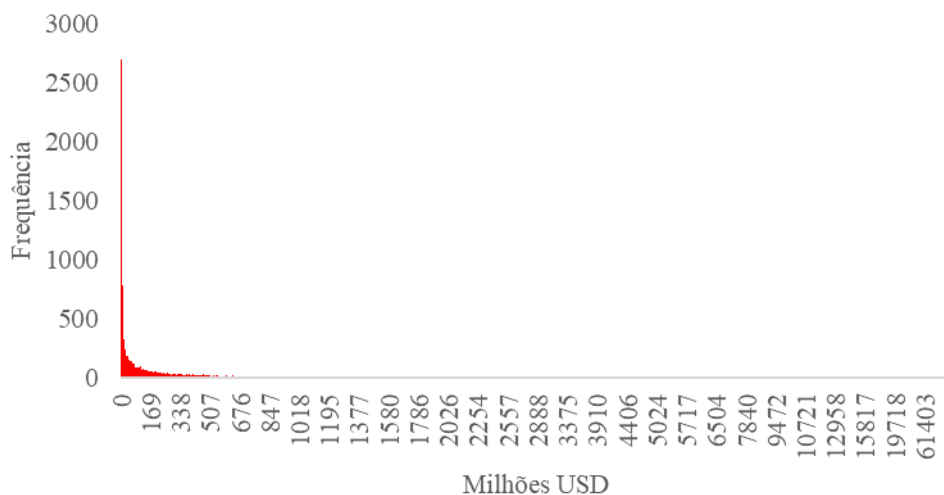


FIGURA 9 – Distribuição dos passivos ao longo do histórico.

As tabelas apresentadas abaixo refletem os valores trimestrais, dos ativos e passivos respetivamente, das estatísticas obtidas, tais como a média, variância acumulada,

excesso de curtose e coeficiente de assimetria. À partida, para poder assumir que uma dada distribuição aproximadamente normal, seria de esperar obter valores para o excesso de curtose inferiores a 6, bem como o coeficiente de assimetria inferior a 2, em valores absolutos. Contudo, não é isso que se verifica.

TABELA I – Valores descriminados das estatísticas obtidas para os valores originais dos ativos.

Data	Média	Variância Acumulada	Curtose	Assimetria
2014T4	15 611 949 024,93	11 913 882 952 331 900 000 000,00	322,04	16,76
2015T1	17 569 387 596,46	14 248 675 866 831 900 000 000,00	339,34	17,21
2015T2	15 841 852 738,49	13 505 574 399 547 500 000 000,00	328,05	16,86
2015T3	16 253 021 265,65	13 126 148 998 611 600 000 000,00	299,00	16,07
2015T4	16 245 436 470,06	12 859 416 583 992 200 000 000,00	299,87	16,09
2016T1	17 880 859 560,54	13 220 919 119 704 600 000 000,00	326,63	16,69
2016T2	17 114 652 311,94	13 259 596 534 048 700 000 000,00	280,55	15,77
2016T3	17 120 517 803,59	13 121 498 823 810 600 000 000,00	301,83	16,05
2016T4	17 899 159 288,13	13 176 963 041 003 300 000 000,00	277,20	15,53
2017T1	19 486 758 259,39	13 735 279 776 137 100 000 000,00	333,87	17,10
2017T2	17 670 511 122,62	13 635 292 464 910 000 000 000,00	306,75	16,11
2017T3	16 823 048 769,96	13 542 296 737 324 700 000 000,00	318,14	16,50
2017T4	16 568 570 404,05	13 433 337 487 370 100 000 000,00	319,16	16,59
2018T1	16 446 346 949,80	13 318 168 860 961 600 000 000,00	325,28	16,76
2018T2	16 204 786 459,58	13 233 493 870 590 200 000 000,00	340,86	17,20
2018T3	16 500 937 363,09	13 163 016 496 810 100 000 000,00	324,04	16,70
2018T4	16 246 527 596,21	13 095 279 667 597 200 000 000,00	342,60	17,24
2019T1	16 083 793 554,48	13 010 343 366 303 700 000 000,00	370,22	17,92
2019T2	16 325 205 625,30	12 957 390 911 308 300 000 000,00	380,05	18,15
2019T3	16 539 550 391,79	12 919 053 407 751 700 000 000,00	375,33	18,06
2019T4	16 810 448 383,70	12 898 377 182 206 600 000 000,00	376,84	18,10

TABELA II - Valores discriminados das estatísticas obtidas para os valores originais dos passivos.

Data	Média	Variância Acumulada	Curtose	Assimetria
2014T4	11 824 292 446,00	8 680 001 599 157 380 000 000,00	361,98	17,84
2015T1	12 716 148 645,00	9 640 192 783 901 030 000 000,00	375,35	18,15
2015T2	11 893 666 664,00	9 302 842 942 888 020 000 000,00	382,86	18,34
2015T3	12 292 115 717,00	9 126 588 860 865 730 000 000,00	341,98	17,31
2015T4	12 300 903 461,00	8 988 395 002 290 010 000 000,00	342,13	17,32
2016T1	12 939 138 033,00	9 067 866 390 110 020 000 000,00	352,01	17,54
2016T2	13 013 001 564,00	9 133 667 517 128 900 000 000,00	309,40	16,67
2016T3	12 909 440 798,00	9 081 864 975 856 290 000 000,00	344,68	17,35
2016T4	13 535 802 508,00	9 144 579 193 002 360 000 000,00	315,82	16,68
2017T1	13 719 969 484,00	9 254 815 752 713 850 000 000,00	358,06	17,65
2017T2	13 371 320 454,00	9 246 139 895 600 750 000 000,00	349,37	17,42
2017T3	12 968 566 406,00	9 237 329 061 295 250 000 000,00	354,91	17,62
2017T4	12 854 925 672,00	9 219 814 462 004 930 000 000,00	361,27	17,82
2018T1	12 800 302 938,00	9 194 660 128 988 940 000 000,00	365,03	17,94
2018T2	12 648 823 248,00	9 189 271 088 611 240 000 000,00	380,90	18,38
2018T3	12 577 767 565,00	9 172 403 243 747 770 000 000,00	379,27	18,33
2018T4	12 704 610 878,00	9 169 012 791 315 280 000 000,00	381,75	18,39
2019T1	12 603 532 206,00	9 152 841 347 306 310 000 000,00	406,20	18,99
2019T2	12 822 047 313,00	9 157 830 803 248 160 000 000,00	417,76	19,25
2019T3	12 975 995 795,00	9 168 921 917 550 750 000 000,00	411,92	19,14
2019T4	13 175 223 301,00	9 191 592 200 451 540 000 000,00	413,61	19,19

A visualização gráfica destes parâmetros pode ser observada abaixo (Figura 10).

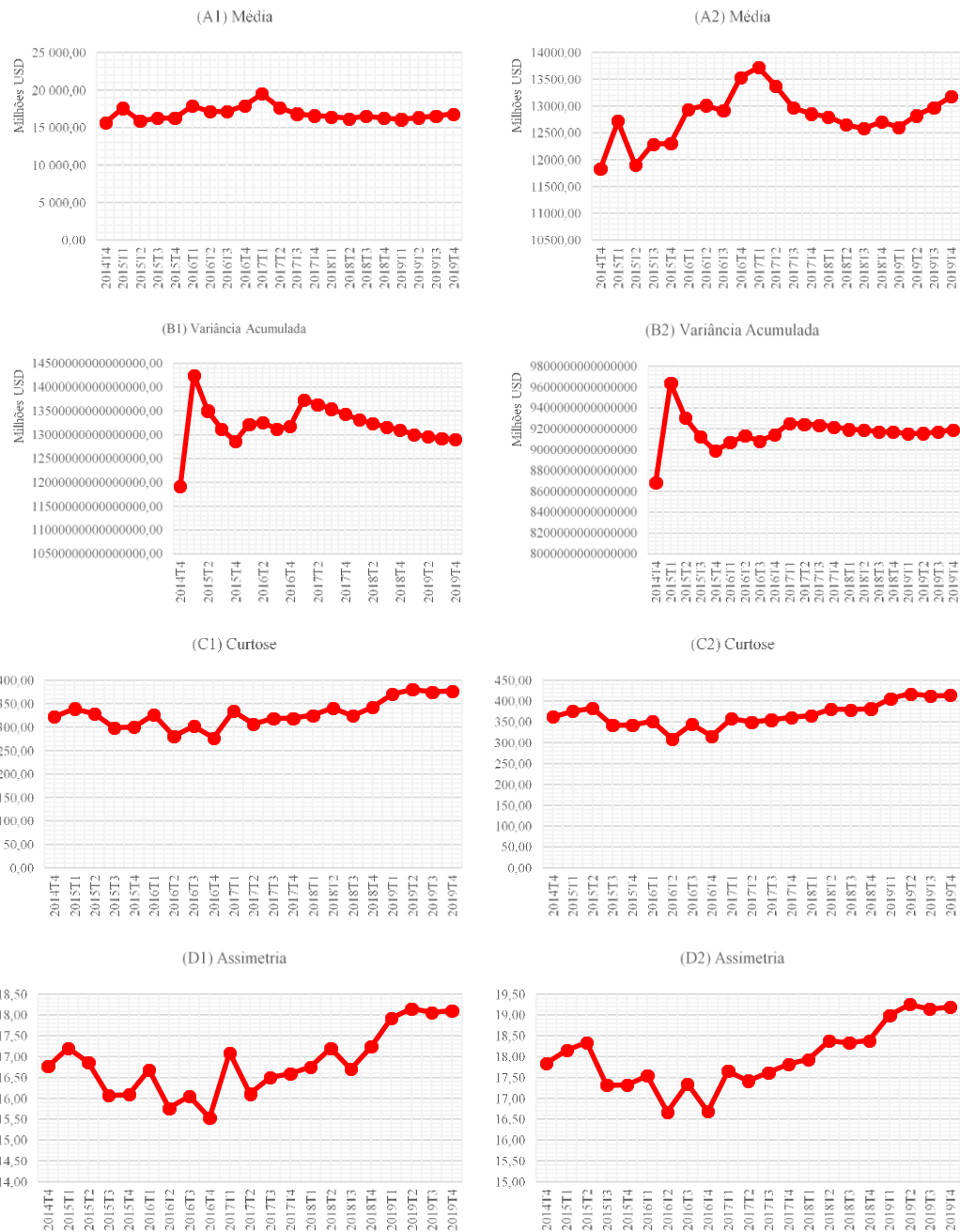


FIGURA 10 - Estatísticas obtidas para ativos (1) e passivos (2) dos dados originais.

De acordo com os resultados obtidos, verifica-se uma pequena estabilidade da média apenas para os ativos. No entanto, quer os ativos quer os passivos apresentam uma variância acumulada divergente, ou seja, não está a diminuir com o tamanho da amostra. Também o coeficiente de curtose apresenta valores muito elevados, indicando que a

distribuição segue uma *fat tail*. Quanto ao coeficiente de assimetria, tudo indica que a existe um enviesamento à direita, tal como foi verificado na Figura 8 e Figura 9.

Após ser aplicado o logaritmo a cada um dos agentes, as tabelas abaixo indicadas refletem os valores trimestrais, dos ativos e passivos respetivamente, das estatísticas obtidas, tais como a média, variância acumulada, excesso de curtose e coeficiente de assimetria.

TABELA III - Valores discriminados das estatísticas obtidas para os valores dos ativos relativizados.

Data	Média	Variância Acumulada	Curtose	Assimetria
2014T4	20,5987	7,6917	9,3545	-1,5751
2015T1	20,6851	7,3865	8,1144	-1,3754
2015T2	20,6755	7,2026	8,5672	-1,4190
2015T3	20,7167	6,9172	7,6454	-1,2551
2015T4	20,7202	7,1815	9,1956	-1,4873
2016T1	20,7715	7,1802	8,1740	-1,3655
2016T2	20,7394	7,1663	8,1343	-1,3562
2016T3	20,7653	7,2973	7,5070	-1,3312
2016T4	20,7910	7,1820	6,8419	-1,2202
2017T1	20,8392	7,2548	7,1021	-1,2561
2017T2	20,8117	7,2119	7,1435	-1,2682
2017T3	20,7990	7,0415	7,4352	-1,3201
2017T4	20,8167	6,9380	7,4229	-1,3123
2018T1	20,8382	6,9777	7,7936	-1,3836
2018T2	20,8393	6,9234	7,8813	-1,3964
2018T3	20,8486	7,0455	8,4036	-1,4857
2018T4	20,8683	7,0190	8,4547	-1,5077
2019T1	20,8667	7,1427	8,1670	-1,5116
2019T2	20,9111	7,0279	8,1985	-1,4974
2019T3	20,9253	7,0596	8,1724	-1,4947
2019T4	20,9403	7,1121	8,0863	-1,4907

TABELA IV - Valores descriminados das estatísticas obtidas para os valores dos passivos relativizados.

	Média	Variância Acumulada	Curtose	Assimetria
2014T1	20,2000	8,0300	2,1800	-0,6400
2015T1	20,2400	8,0700	2,3700	-0,6700
2015T2	20,2200	8,1800	2,3300	-0,6800
2015T3	20,2500	8,1600	1,4500	-0,5600
2015T4	20,2800	8,1300	2,4000	-0,7300
2016T1	20,3100	8,1300	1,5900	-0,6000
2016T2	20,3200	8,1300	1,5900	-0,6300
2016T3	20,3500	8,1200	1,5700	-0,6200
2016T4	20,3600	8,1300	1,6200	-0,6300
2017T1	20,3900	8,1300	1,6700	-0,6300
2017T2	20,3700	8,1400	1,6300	-0,6500
2017T3	20,3700	8,1400	1,7700	-0,6800
2017T4	20,3700	8,1400	1,8300	-0,6900
2018T1	20,4000	8,1400	1,8800	-0,7100
2018T2	20,3900	8,1400	1,7700	-0,7100
2018T3	20,4000	8,1500	1,8500	-0,7300
2018T4	20,4200	8,1600	2,1700	-0,8100
2019T1	20,4200	8,1700	1,8700	-0,7600
2019T2	20,5000	8,1600	2,0300	-0,8000
2019T3	20,5100	8,1600	2,0600	-0,8000
2019T4	20,5500	8,1500	2,0100	-0,7900

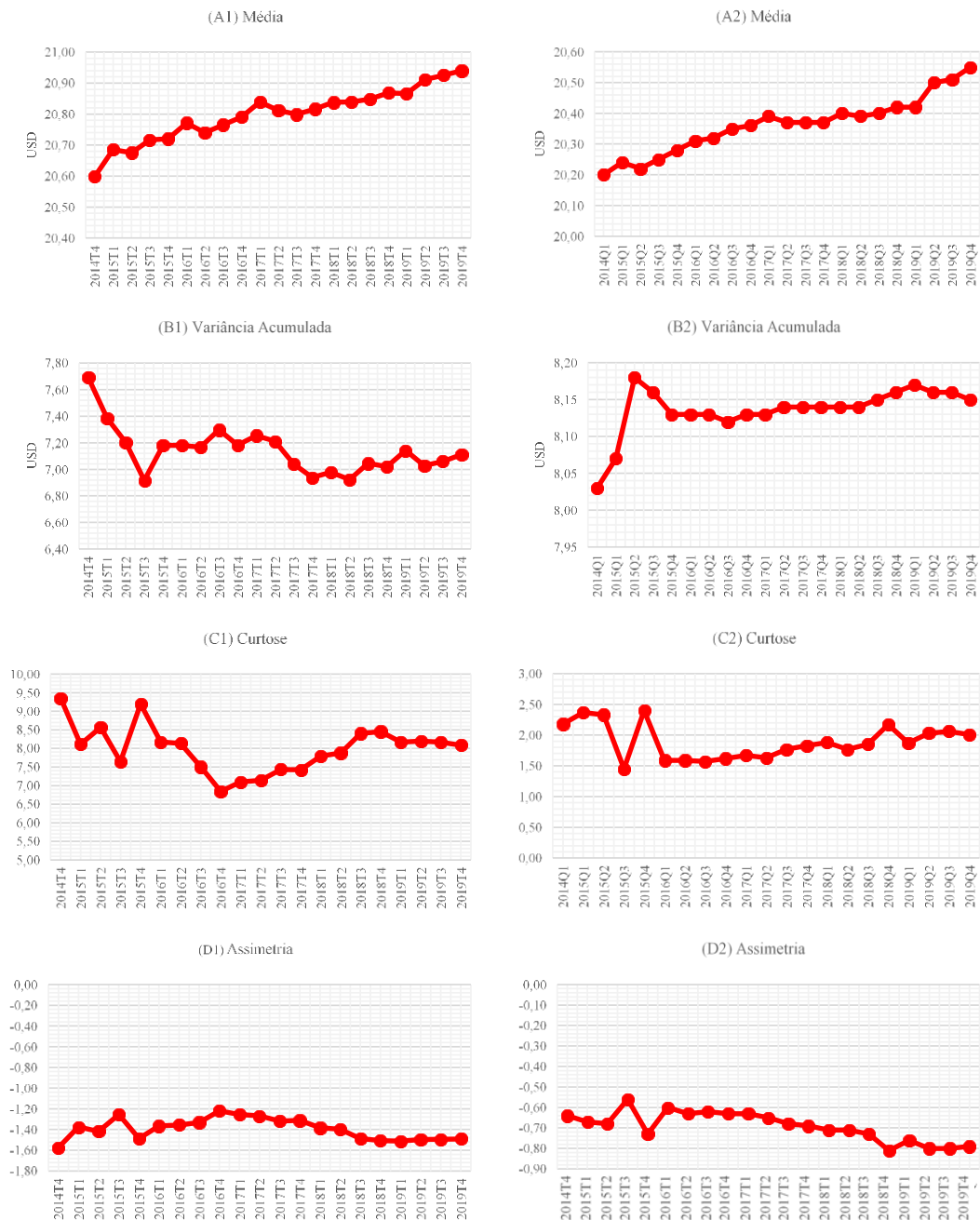


FIGURA 11 - Estatísticas obtidas para ativos (1) e passivos (2) com dados relativizados.

Aquando é calculado o logaritmo para cada um dos agentes, observa-se uma variância acumulada não convergente quer para os ativos quer para os passivos. Também os coeficientes de excesso de curtose e assimetria assumem valores mais baixos. Ainda assim, é possível observar uma diferença nos valores entre os ativos e passivos,

nomeadamente no coeficiente de curtose. Contudo, podemos afirmar que após esta modelação verificamos os valores esperados para a obtenção de uma distribuição normal (ou quase normal, neste caso), uma vez que a curtose é inferior a 6 e o valor absoluto do coeficiente de assimetria é inferior a 2. No entanto, o valor da média está a crescer em ambos os casos.

Após a aplicação do algoritmo, as tabelas apresentadas abaixo refletem os valores trimestrais, dos ativos e passivos respetivamente, das estatísticas obtidas, tais como a média, variância acumulada, excesso de curtose e coeficiente de assimetria.

TABELA V - Valores discriminados das estatísticas obtidas para os valores dos ativos após a aplicação do algoritmo.

Data	Média	Variância Acumulada	Curtose	Assimetria
2014T4	0,0000	7,6854	9,3522	-1,5748
2015T1	0,0003	7,5363	8,1321	-1,3792
2015T2	-0,0002	7,4244	8,5844	-1,4223
2015T3	-0,0006	7,2979	7,6133	-1,2519
2015T4	-0,0007	7,2729	9,1716	-1,4856
2016T1	-0,0002	7,2574	8,2102	-1,3694
2016T2	0,0001	7,2435	8,0890	-1,3518
2016T3	0,0001	7,2497	7,5453	-1,3346
2016T4	0,0004	7,2419	6,8540	-1,2214
2017T1	-0,0003	7,2426	7,0680	-1,2521
2017T2	0,0004	7,2393	7,1191	-1,2659
2017T3	0,0001	7,2226	7,4330	-1,3188
2017T4	0,0000	7,2006	7,3959	-1,3085
2018T1	-0,0003	7,1841	7,7726	-1,3828
2018T2	-0,0009	7,1667	7,8333	-1,3924
2018T3	0,0011	7,1589	8,3614	-1,4830
2018T4	0,0013	7,1507	8,4646	-1,5088
2019T1	-0,0006	7,1503	8,1934	-1,5142
2019T2	0,0015	7,1439	8,1644	-1,4925
2019T3	0,0006	7,1397	8,1353	-1,4897
2019T4	0,0008	7,1382	8,0453	-1,4873

TABELA VI - Valores descriminados das estatísticas obtidas para os valores dos passivos após a aplicação do algoritmo.

Data	Média	Variância Acumulada	Curtose	Assimetria
2014T1	-0,3800	9,4400	5,0800	-1,1000
2015T1	-0,3600	9,3300	4,7300	-1,0400
2015T2	-0,3300	9,2900	4,7300	-1,0400
2015T3	-0,3400	9,2400	4,0800	-0,9500
2015T4	-0,3500	9,2500	4,7900	-1,0900
2016T1	-0,3300	9,2300	4,2000	-0,9900
2016T2	-0,3500	9,2300	4,1000	-1,0000
2016T3	-0,3300	9,2300	4,1000	-1,0000
2016T4	-0,3300	9,2300	4,2000	-1,0100
2017T1	-0,3100	9,2200	4,3500	-1,0200
2017T2	-0,3100	9,2200	4,2000	-1,0300
2017T3	-0,3000	9,2100	4,4600	-1,0700
2017T4	-0,2800	9,1900	4,5700	-1,0800
2018T1	-0,2800	9,1800	4,6400	-1,1100
2018T2	-0,2800	9,1700	4,5200	-1,1000
2018T3	-0,2700	9,1500	4,6000	-1,1300
2018T4	-0,2800	9,1500	4,7500	-1,1800
2019T1	-0,2700	9,1500	4,5800	-1,1400
2019T2	-0,2700	9,1400	4,8500	-1,2000
2019T3	-0,2600	9,1200	4,9000	-1,2000
2019T4	-0,2700	9,1100	4,8800	-1,1900

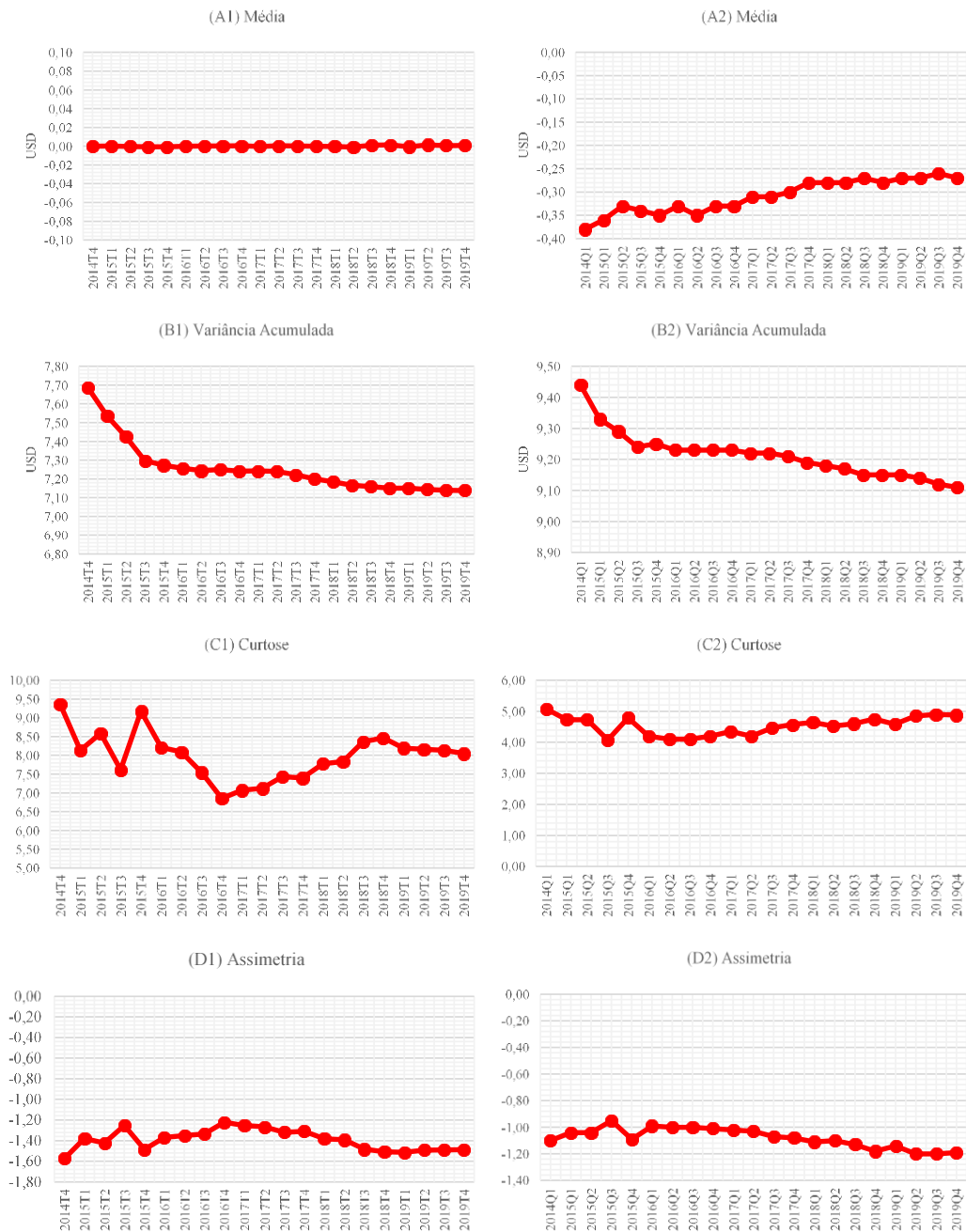


FIGURA 12 - Estatísticas obtidas para ativos (1) e passivos (2) depois da aplicação do algoritmo.

Após ser aplicado o algoritmo de transformação, é possível afirmar que a média para os ativos apresenta um comportamento estável, o que não se verifica para os passivos, contudo a média tende a estabilizar nos anos mais recentes. Já a variância acumulada diminui com o aumento da amostra, como era esperado para termos um erro

menor com o tamanho da amostra. No entanto, para os valores obtidos para o excesso de curtose, verifica-se que no caso dos ativos obtemos valores superiores daqueles que se verificam nos passivos, este abaixo do valor esperado para uma distribuição exponencial (6). No caso dos ativos está acima desse valor e é inconclusivo se com mais dados pode vir a descer.

Após a aplicação do algoritmo desenvolvido, para todo o histórico obtemos a seguinte distribuição para ativos e passivos:

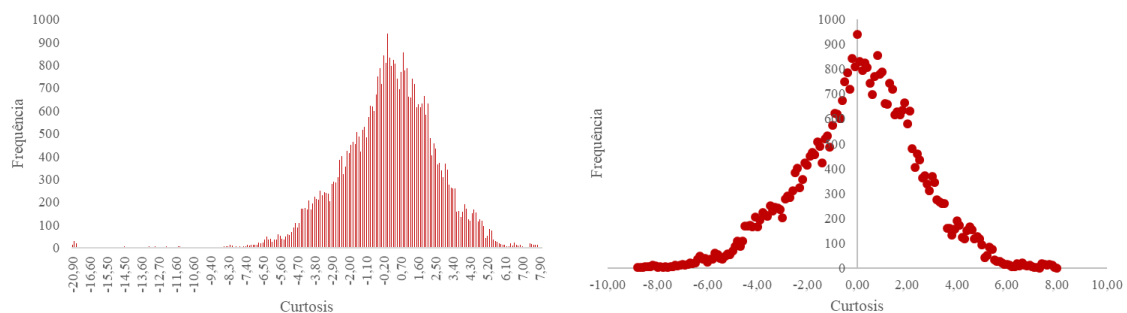


FIGURA 13 – Distribuição dos ativos após a aplicação do algoritmo.

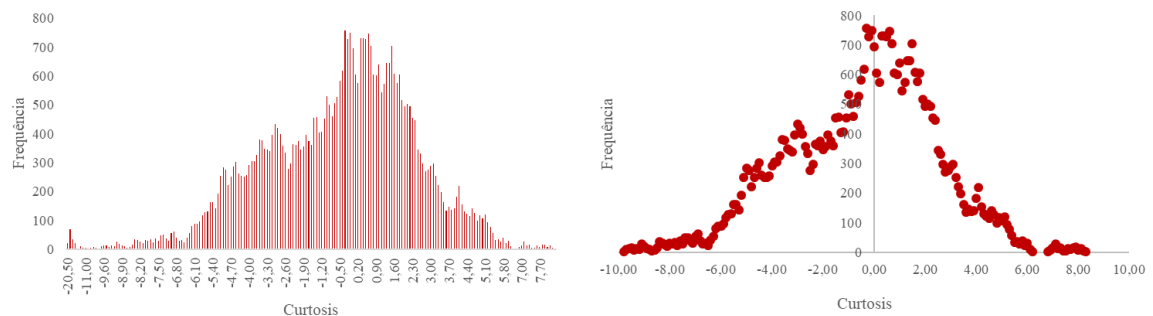


FIGURA 14 – Distribuição dos passivos após a aplicação do algoritmo.

De acordo com os resultados acima apresentados, verificamos uma distribuição próxima da exponencial quer para os ativos quer para passivos, uma vez que se verificam os pressupostos de uma distribuição normal ($\text{curtose} < 6$ e $|\text{assimetria}| < 2$)

6. CONCLUSÃO

O presente documento propõe um algoritmo de normalização da distribuição de uma variável económica. De uma forma simplista podemos escrever que os dois principais desafios observados são a não Gaussinidade que estas grandezas apresentam e a constante expansão do espaço onde se representam. Para ultrapassar ambos foi necessário recorrer a duas áreas de conhecimento: o estudo estatístico de variáveis económicas e a teoria de grafos. O primeiro permite uma normalização após a qual o TLC é aplicável; o segundo é necessário para integrar a expansão do espaço económico na normalização.

É importante realçar que este algoritmo tem duas limitações cujas consequências podem ser exploradas recorrendo a outras técnicas: uma amostra normalizada continua a não ser normal. Apesar da variância ser finita, ainda não é claro quão válido é assumir que a distribuição terá uma média e desvio padrão bem definidos; por outro lado o estudo da expansão do universo é muito dependente da qualidade da informação registada. Ainda no âmbito da qualidade da informação, é importante deixar uma ressalva na extração e tratamento da informação, existindo a possibilidade de ter ocorrido algum erro em algum dos processos, levando, por isso, a um enviesamento dos dados apresentados

Ao longo deste trabalho foi tido um especial cuidado na ligação entre a Física e a Economia. Ao progredir numa direção específica (normalizando as distribuições com variância infinita) procurou-se estabelecer uma ligação sólida entre as diversas áreas de conhecimento usada juntamente com a bibliografia que ajude a dar os primeiros passos neste cruzamento entre física, economia e estatística.

Um campo onde a estratégia proposta pode encontrar aplicações é o de *Machine Learning*. Um conceito fundamental, que é muitas vezes desprezado, nesta área do conhecimento é o quão aprendível é uma população. A aplicação de algoritmos a dados desadequados levará a conclusões erradas. Assim, estes algoritmos podem integrar o pré-tratamento de dados económicos para, por exemplo, se cada registo tiver associada uma etiqueta de vivo/falido poderá ajudar a perceber qual o rácio ideal de ativo/passivo para que uma empresa não vá à falência. Também a capacidade de incorporar comportamentos

realistas poderá ser explorada ainda mais com o recurso da inteligência artificial. Num futuro não tão longínquo, estas técnicas poderão ser usadas para “treinar” um agente artificial para se comportar como um consumidor ou empresa. Aos avanços na inteligência artificial podem tornar os modelos baseados em agentes mais realistas e, quem sabe, prever comportamentos emergentes.

REFERÊNCIAS BIBLIOGRÁFICAS

- Barabási, A.-L., & Réka, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*. 74, 47-97.
- Biggs, N., Lloyd, E., & Wilson, R. (1986). *Graph Theory*. Oxford: Clarendon Press.
- Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81.3, 637-654.
- Blanchett, T., Fournier, J., & Piketty, T. (2017). Generalized Pareto Curves: Theory and Applications. *WID.world Working Paper Series 2017/3*.
- Borland, L., Bouchaud, J.-P., Muzy, J.-F., & Zumbach, G. (March de 2005). The Dynamics of Financial Markets - Mandelbrot's Multifractal Cascades, and Beyond. *Wilmott Magazine*, 86-96.
- da Cruz, J., & Lind, P. (2012). The dynamics of financial stability in complex networks. *The European Physics Journal B: Condensed Matter and Complex Systems*, 85(8), 1-9.
- da Cruz, J., & Lind, P. G. (2013). The bounds of heavy-tailed return distributions in evolving complex networks. *Physics Letters A, Vol.377*, 3-4, 189-194.
- Degli Atti, M., Meler, S., Rizzo, C., Ajelli, M., Massari, M., Manfredi, P., Iannelli, M. (2008). Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios. *PLoS ONE*, 3(3), e1790, 1-11.
- Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Euler Archive - All Works*. 53, 128-140.
- Gnedenko, B., & Kolmogorov, A. (1955). Limit Distributions for Sums of Independent Random Variables. *Mathematica Scandinavica Vol.3, No. 1*, 185-187.
- Gualdi, S., Tarzia, M., Zamponi, F., & Bouchaud, J. (2015). Tipping points in macroeconomic agent-based models. *Journal of Economic Dynamics & Control* 50, 29-61.
- Hoag, A., & Hoag, J. (2006). *Introductory Economics*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Honerkamp, J. (2012). *Statistical Physics - An Advanced Approach with Applications*. New York: Springer.
- Hong, H., & Stein, J. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 2, 43-84.

- Kampen, N. G. (2006). *Stochastic Processes in Physics and Chemistry*. Amsterdam: Elsevier.
- Mantegna, R. N., & Stanley, H. (1999). *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge: Cambridge University Press.
- Mantegna, R. N., & Stanley, H. E. (1994). Stochastic Process with Ultraslow Convergence to a Gaussian: The Truncated Lévy Flight. *Physical Review Letters* 73(22), 2946-2949.
- Mirowski, P. (1989). *More Heat than Light: Economics as Social Physics, Physics as Nature's Economics*. Cambridge: Cambridge University Press.
- Pareto, V. (1964). *Cours d'Économie Politique*. Geneva: Librairie Droz.
- Pires, J., Cruz, H., Rajaratnam, K., Beling, P., & Overstreet Jr., G. A. (2016). Sand Pile Modeling of Economic Variables For Credit Risk Applications. *CSCC XIV*.
- Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science*. London: Macmillan and Co., Limited.
- SEC.GOV. (s.d.). Obtido em 6 de janeiro de 2020, de U.S. Securities and Exchange Commission: <https://www.sec.gov/dera/data/financial-statement-data-sets.html>
- Shoutens, W. (2003). *Lévy Processes in Finance: Pricing Financial Derivatives*. Chichester: Wiley.
- Smith, E., & Foley, D. K. (2008). Classical thermodynamics and economic general equilibrium theory. *Journal of Economic Dynamics and Control* 32.1, 7-65. doi:10.1016/j.jedc.2007.01.020
- Sociedade - Rede Complexa de Relações. (2020). Obtido de Instituto Blaise Pascal - Tecnologia e Educação: <http://www.institutopascal.org.br/canais/organizacoes/organizacoes.php>
- Sornette, D., & Cont, R. (1997). Convergent Multiplicative Processes Repelled from Zero: Power Laws and Truncated Power Laws. *Journal de Physique I, EDP Science* 7(3), 431443.
- Turrell, A. (Q4 de 2016). Agent-based models: understanding the economy from the bottom up. *Quarterly Bulletin - Bank of England*, pp. 173-188.

ANEXOS

TABELA VII – Base de dados utilizada na modelação¹.



Base de Dados.xlsb

¹ Em alternativa, o ficheiro pode ser consultado em: https://phdisegutl-my.sharepoint.com/:x/g/personal/cristinacf_aln_iseg_ulisboa_pt/EZUulfz9iXNNkEFZgxvjRlsBiLZ5KgFvU91iS3H29EjRAg?e=sVdz2r